生成 AI が拓くデータ活用の新境地

Google Cloud の 「データ エージェント」とは？

Google Engineers

山田 雄 | Yu Yamada

Customer Engineer
Data Analytics Specialist

Google Cloud
Innovators

- データ分析専門のカスタマーエンジニア
- 15年以上データ分析基盤の開発・運用に従事
- 日本最大級のデータ エンジニア コミュニティ主催
  - datatech-jp
- 好きなプロダクトは **BigQuery** と **Dataplex**

著書

レビュー

データ活用の「今」と「未来」	01
Google Cloud の「データ エージェント」とは？	02
「データ エージェント」がもたらすビジネス変革のシナリオ	03
データ駆動型組織への新たな一歩	04

データ活用の「今」と「未来」	01
Google Cloud の「データ エージェント」とは？	02
「データ エージェント」がもたらすビジネス変革のシナリオ	03
データ駆動型組織への新たな一歩	04

The Data Value

raw data から実用的な価値のあるデータへ： 速度と複雑性の問題

手作業では時間がかかる

従来の ETL は手作業でコーディングに時間がかかり時間を消費する

イノベーションが起きづらい

新しいデータを使った実験や競合優位のための高度な分析を迅速に試すことが困難

サイロとマルチモーダル

様々なシステムをつなぎ合わせた分析や、様々なフォーマットの分析が困難

データを取り扱う上での課題

専門家でないデータから価値を引き出しにくい
データの民主化が進まない

Autonomous Data and AI platform

Powered by **Gemini**

“サイロ化されたインサイト” から “統合インテリジェンス” へ

BigQuery

データの取り込みから AI による分析まで、データのライフサイクル全体を支援して、手作業の労力を最小限に抑える

BigQuery と Vertex AI のシームレスな統合により、ML/AI モデルのライフサイクルを迅速に加速し、多様なチームをエンパワー

自律的な data-to-AI platform.

複雑なタスクを自動化し、問題を積極的に特定して解決し、専門のアナリスト、エンジニア、保護者として機能する AI エージェント

マルチモーダル データ、AI 駆動型分析、セマンティック インテリジェンス、LLM コプロセッシング、多様なエンジン、会話型 BI、リアルタイム機能を統合した統合プラットフォーム

データ活用の「今」と「未来」	01
Google Cloud の「データ エージェント」とは？	02
「データ エージェント」がもたらすビジネス変革のシナリオ	03
データ駆動型組織への新たな一歩	04

AI Agentとは？

特定の目的を達成するために、環境を認識し、状況を判断し、自律的に行動を実行する AI システム。

以下のような要素を持ち合わせている

認識 (Perception) : まるで人間の目や耳のように、周囲の状況や情報(例:ユーザーからの指示、アクセス可能なデータ、他のシステムからの情報)を理解する能力。

推論・計画 (Reasoning & Planning) : 認識した情報に基づいて、目標達成のための最適な方法を考え、どのように行動すべきか、具体的な計画を立てる能力。これは、人間が「どうすればうまくいくか」を考えるのと同じようなプロセスです。

行動 (Action) : 計画に従って、実際にシステム内で具体的なタスクを実行する能力。例えば、データ分析の実行、レポートの作成、他システムへの連携などです。

記憶・学習 (Memory & Learning) : 過去の対話や実行結果を記憶し、そこから学び、次回のパフォーマンスを向上させる能力。これにより、使えば使うほど賢く、パーソナライズされていきます。

自律性 (Autonomy) : そして、これが最も重要な点ですが、人間の細かい指示がなくても、自ら状況を判断し、複数のタスクを連携・実行できる能力を指します。

What is an 'AI agent'?

入力情報とツールを活用し、目的のための最適な方法を推論するアプリケーション

Key components

Model:

目標を推論し、計画を決定し、レスポンスを生成するために使用されます。

Tools:

他のAPIやサービス呼び出すことで、データを取得したり、アクションやトランザクションを実行したりします。

Orchestration:

メモリと状態（計画に使用されたアプローチを含む）、ツール、提供されたデータや取得されたデータなどを維持・管理します。

Agent

Building blocks of an agent

- Model **The Brain**
- Tools **The Hands**
- Orchestration **The Coordinator**
- Memory/ Context / History

Agents in action: The power of tools

エージェントが利用するツールは一つだけではありません。常に業務に合わせて、最も適切なツールを選択します

データ分析からプロセス自動化まで、強力な機能を活用可能に

シームレスに連携する、単一の統合型ツールプラットフォーム

From individuals to teams: The multi-agent approach

専門スキルを持つエージェントたちがチームを組み、互いの知見を共有しながら連携して、大きな課題を解決

Data Agents

データ エージェント ファミリー

Data Engineering Agent

データ準備、データ取り込み、データクリーニング、ブロンズ→シルバー→ゴールドの自動化

Data Governance Agent

データモデリング、メトリック定義、メタデータ生成

Data Science Agent

Colab ノートブックエージェント、探索的データ分析 (EDA)、データ可視化、予測モデリング

Conversational Analytics Agent

Data QnA, フォークキャスト, マルチモーダル検索

Data Engineering Agent

自動化されたデータ パイプラインの
作成、キュレーション、最適化。
BigQuery とのシームレスな連携

- データ パイプラインを大幅に加速
- 全てのデータ関連ステークホルダーを支援
- データ品質とガバナンスを保証
- データ エンジニアリング コストを削減
- ビジネス インサイトを即座に解き放つ

データ エンジニアの工数削減に

BigQuery Data Preparation

BigQuery Studio 内のデータプリパレーションエージェント

- Gemini の支援により
SQL データパイプラインを数分で開発
- コンテキストを考慮した
AI 生成の変換機能により、データの検出、
クレンジング、変換を実行可能に
- スキーマ マッチング、
ビジネスとデータ品質のアサーション生成を
自動化することで、時間を節約
- GitHub との連携により、BigQuery で
デプロイとオーケストレーションを実行可能

Data Preparation / 注文データをクレンジング

The screenshot shows a data preparation tool interface. On the left, a table displays order data with columns for 'purchased_date' and 'prefecture'. The 'purchased_date' column is highlighted, and a bar chart icon is visible. The table contains 12 rows of data. On the right, a 'ステップ' (Steps) panel shows a suggestion from Gemini AI to convert the 'purchased_date' column from a string to a date using a specific format. The suggestion includes a preview button and an '適用' (Apply) button.

	purchased_date	prefecture
1	09/30/2024	愛知県-名古屋市
2	01/10/2024	愛知県-名古屋市
3	01/10/2024	滋賀県-大津市
4	09/30/2024	栃木県-宇都宮市
5	2024_09_30	山梨県-甲府市
6	2024_10_01	神奈川県-横浜市
7	2024_09_30	三重県-津市
8	09/30/2024	埼玉県-さいたま市
9	2024_09_30	東京都-東京
10	2024_09_30	茨城県-水戸市
11	2024_10_01	滋賀県-大津市
12	2024_10_01	愛知県-名古屋市

ステップ

提案 (10) 適用されたステップ (1)

✦ Gemini による提案

AI による分析情報を活用してデータを準備します。

変換

Converts column purchased_date from type string to date with the following format(s): %m/%d/%Y;%Y_%m_%d

```
COALESCE(SAFE.PARSE_DATE("%m/%d/%Y", purchased_date), SAFE.PARSE_DATE("%Y_%m_%d", purchased_date))
```

この提案を評価してください

プレビュー 適用

+ ステップを追加

Gemini in BigQuery - AI for everyone

Gemini の主要機能が無料で利用可能に

- SQL コード生成 / 説明 / 補完
- Python コード生成 / コード補完
- Data canvas
- Data preparation

The screenshot shows the Google Cloud Data Preparation interface. The main area displays a table with columns: channel, impressions, clicks, bookings, campaign_date, local_temp, hotel_name, and phone. The table contains data for various hotels and campaigns. On the right side, there are 'Suggestions by Gemini' and 'Standardize phone numbers' sections, each with a 'Filter' button and a 'Suggest more' button. The 'Filter' section shows a filter rule: 'NOT regexp_match(zipcode, '[0-9]{3}-[0-9]{3})\$)'. The 'Standardize phone numbers' section shows a rule: 'PREP_NORMALIZE_PHONE('US', phone)'. The interface also includes a 'Steps' panel on the right and a 'Details' panel at the bottom.

channel	impressions	clicks	bookings	campaign_date	local_temp	hotel_name	phone
string	integer	integer	integer	string	string	string	string
social_media	560000	12500	485	15-07-2013	82°F	The Liberty Inn	(212) 555-1212
email	120000	8700	230	22-12-2023	28°C	Outback Oasis Resort	+61 2 9876 5432
social_media	560000	12500	485	15-07-2013	82°F	Redwood Lodge and Cabins	1-707-555-9800
email	120000	8700	230	22-12-2023	28°C	Wallaby Creek B&B	02 6123 4567
display_ads	340000	4200	175	09-08-2023	26°C		
partner_site	85000	9500	350	28-06-2023	31°C	Ryokan Sakura	03-5555-2222
search_ads	410000	18500	620	10-11-2023	-5°C	Maple Leaf Lodge	(416) 555-3210
social_media	280000	11300	315	14-02-2023	9°C	The Crown and Sceptre	020 7948 0055
display_ads	610000	8600	540	28-07-2023	33°C	Hacienda del Sol	+52 (624) 143-5555
email	220000	6300	290	25-09-2023	25°C	Pousada da Floresta	(11) 4002 8922
partner_site	170000	14100	410	27-12-2023	30°C		
email	92000	7200	240	18-03-2023	12°C	Kyoto Tea House Inn	(075) 351-2345
search_ads	350000	22700	875	20-08-2023	84.2°F	The Jazz Corner Hotel	+1-504-555-9080
social_media	420000	6400	280	05-07-2023	30°C	Villa Toscana	+39 055 555566
display_ads	375000	5800	210	01-12-2023	4°C	Gasthaus zum Adler	(089) 22 33 44
partner_site	670000	30000	940	12-02-2023	28°C	The Taj View Hotel	91-562-222-1234
search_ads	450000	16400	620	15-08-2023	32°C		
email	140000	6700	185	28-12-2023	22°C	Kiwi Country Lodge	+64 3 543 9876
social_media	380000	20100	790	25-08-2023	24°C	Hockey Haven Hotel	1-404-HOCKEY
display_ads	385000	5300	190	20-07-2023	35°C	Santorini Sunset Villas	30-22860 22548
partner_site	110000	10200	360	05-08-2023	68°F	Crossroads Motel	505-555-1234
search_ads	290000	13700	430	20-06-2023	10°C	Fincas Los Gauchos	+54-11-5555-7777
social_media	420000	16100	515	15-02-2023	25°C	Playa del Carmen Paradise	(984) 876 5432

高度なコンテキストとプロアクティブな管理ツールを柔軟な価格帯から選択

BigQuery Pipeline

グラフィカルなパイプラインをBigQuery上で構成

- ワークフローで SQL スクリプト、Python ノートブック、Data Preparation を定期実行可能
- ノーコードでパイプラインを視覚的に構築
- BigQuery Studio に統合
BigQuery のすべてのスケジュールジョブを単一画面から監視
- エンジニアに頼らずワークフローの作成が可能

Explorer

+ ADD

ELT daily

Type to search

- project-dev
  - Queries
  - Notebooks
    - my notebook
  - Pipelines
    - ELT daily**
  - Repositories
    - my repository
  - stations-data
  - transactions
    - tx
    - tx_labels

ELT daily

Run Publish Schedule Ask Agent

Pipeline Executions Settings

Add task 100%

Let's try 'Ask Agent', our AI-powered data engineering agent that automates and simplifies the creation, operation and maintenance of pipelines.

Data Governance Agent

データ ガバナンス タスクを自動化

- **Proactive Governance:** コンプライアンス上の不備が発生するのを待つのではなく、システムがデータ分類、アクセス制御、データリネージ追跡、ポリシー更新を積極的に提案
- **Data enrichment:** テーブルや列のメタデータを自動的に生成し、拡充
- **Semantic grounding for agents:** エージェントにとって、統制の取れた信頼できる唯一の情報源 (SoT: Source of Truth)
- **Knowledge Graph-Powered Insights:** ナレッジグラフを利用してデータセット、メタデータ、ポリシー間の関連性を解明(マッピング)し、リアルタイムでガバナンスに関する推奨事項を提供

Automated metadata curation

- 自動メタデータ キュレーションは、明確なドキュメントの欠如や、エラーが発生しやすい手動メタデータ作成に起因する課題に対処します。特にデータセットの規模と複雑さが増すにつれて、こうした課題が顕在化します
- BigQuery データ インサイト スキャンは、**テーブルと列の説明**、そしてクエリの推奨事項を**自動生成**します
- ユーザーは必要に応じて説明を確認、編集し、スキーマタブと詳細タブにコピーできます

The screenshot displays the Google Cloud BigQuery interface. On the left, the Explorer pane shows a project named 'insightshughash' with a table 'spotify_2023' selected. The main pane shows the 'Table description' for 'spotify_2023', which includes a table description and a list of column descriptions. The table description states: 'This table contains information about Spotify tracks, including their audio features and popularity across different music platforms. It provides insights into track characteristics such as tempo, energy, and instrumentality. The table also offers data on track presence in charts and playlists on platforms like Apple Music, Deezer, and Spotify. This allows for analysis of track popularity and performance across various music platforms.' Below the table description, there are several generated insights, such as 'Identify artists with an unusually high number of tracks in Spotify charts compared to their average track presence in playlists.' and 'Find tracks with exceptionally high 'iveness' scores, indicating potential live recordings, that fall outside the typical range for their genre (defined by 'mode)'. The interface also shows a 'Filter' section and a 'Job history' section.

自動的なデータ異常探知

データの信頼性を向上

BigQuery の Data Quality 機能で、BigQuery テーブルとビューのデータの問題を監視して自動的に検出する自動データ異常検出のサポートを開始

能動的なリスク対処が可能に

異常を特定し、コストのかかるミス、不正行為、コンプライアンス違反がビジネスに影響する前に防止

AI の力を解き放つ

高度な機械学習アルゴリズムを活用して、従来のルールベースのシステムでは見逃される可能性のあるパターンや異常を自動的に検出

Data Science Agent

データサイエンスのタスクを自動化する 1st Party エージェント

以下の作業を自動化：

- **データ ラングリング**：データフレームの自動取り込みとクリーニング
- **データ探索**：ネイティブの視覚化と自動探索分析
- **AIによる予測**：Google Research の TimesFM などの最先端のモデルを含む AI を活用した予測
- **モデル評価**：精度と生産準備状況をより深く理解するためのモデル評価

Callout Agent session

プレビュー

Conversational Analytics Agent

データ分析タスクを自動化する
1st Party エージェント

以下の作業を自動化：

- **分析の民主化**：自然言語ベースのインサイト
- **Custom data agents**: 迅速な導入を可能にするクイックスタートを備えた柔軟な API
- **Direct connect to BigQuery**: BigQuery でデータを直接探索するか、Looker のセマンティック レイヤーと組み合わせる
- **Python の Code interpreter**: 複雑なデータ サイエンス タスク用の Python コードを生成

Conversational Analytics API

エンタープライズ横断で Data Agent をカスタマイズ & デプロイ

Cymbal Pets Dashboards Chat Team Help Log out

← Talk to data + New conversation

What if Las Vegas customers purchase 20% more this week? What will be the impact on our supply chain and sales forecast?

Sounds like a good plan, let's get into it! First let's see what current combined numbers are for x, y and z...

Date	Supply	Sales
1/15/25	124	5.2M
1/14/25	124	3.9M
1/13/25	135	3.9M
1/12/25	69	3.3M
1/11/25	45	3.1M
1/10/25	122	2.2M
1/9/25	100	1.6M

BigQuery

Looker Model

カスタム エージェントを迅速に開発

クイックスタートを備えた
柔軟な API で迅速な導入を実現

分析目的に特化した API

ビジネス インテリジェンス、データ探索、
分析向けにチューニングされた AI サービス

BigQuery へ直接接続可能

BigQuery でデータを直接探索するか、
Looker のセマンティックレイヤと組み合わせる

再利用可能でスケーラブルな セマンティック レイヤー

精度の向上と高度なマルチテーブル分析

Google Cloud コンソールに入らずに
Slack など普段使い慣れた IF からのデータ分析が可能

Looker Data Barista

Chat Home About

Now

Whats the average sales for first time visitors

Looker Data Barista App Now

Average sales for first time visitors is **4.30 USD**. This only includes registered customers!

Now

what questions can you help me with?

Looker Data Barista App Now

This datasource contains information about the state of our Looker Coffee business. Here are some example questions you could ask:

- How many total orders did we serve?
- What is our customers favorite product?
- How are our locations in Berlin performing?

You can also share images with me along with a question.

History is on

- 2
- 1
- 10
- 1

- 31
- +

BigQuery Knowledge Engine

- BigQuery Knowledge Engine は、**Knowledge Graph** の力を活用して、高度なデータインサイトとデータ連携を実現します。
- 多様なデータセットとメタデータを統合することで、ユーザーはテーブル間の**隠れた関係性**を発見し、テーブル間クエリを生成し、より深い分析インサイトを得ることができます

Google Cloud Knowledge Engine interface showing a Knowledge Graph and Relationships table.

Knowledge graph generated by Gemini

Dataset description

This dataset contains details about customers of a telecommunications company, including their subscription plans, service usage, and account information. Data exploration should focus on identifying patterns and factors that contribute to customer churn, with the goal of improving customer retention strategies.

Knowledge graph

Relationships

Table 1	Table 2	Relationship
☐ Customers	CustomerProfiles	Customers.customerID = CustomerProfiles.customerID
☐ Customers	Orders	Customers.customerID = Orders.customerID
☒ Customers	CustomerProfiles	Customers.id = CustomerProfiles.customerID
☐ Customers	Orders	customers.ID = Orders.customer_ID AND c...
☐ Transactions	BillingPeriods	EXTRACT(YEAR FROM t.TransactionDate) = ...

Insights

Insights generated by Gemini

What are the top 5 most popular program genres and how many times were they accessed?

Find the total number of executed cases for each product

What are the top 5 most popular program genres and how many times were they accessed?

MCP サーバと独自 Agent

社内ツールなどから独自 Agent を使い MCP を経由して BigQuery のデータ操作も可能
セキュアな環境にて構築がしやすい

MCP サーバと自前 Agent

Discover, build and deploy agents

データ活用の「今」と「未来」	01
Google Cloud の「データ エージェント」とは？	02
「データ エージェント」がもたらすビジネス変革のシナリオ	03
データ駆動型組織への新たな一歩	04

メダリオン アーキテクチャ

メダリオン アーキテクチャ with Data Agents

The Data Value

The Data Value

部門・役割ごとの未来像

部門・役割ごとの未来像:

- 営業担当者: 「この地域の最新の売上トレンドと、成功しているキャンペーンは？」といった質問をデータエージェントに投げかけ、リアルタイムで顧客データや市場動向を分析し、商談に活かす。
- マーケティング担当者: 「最新のキャンペーンの ROI を前年と比較して教えて。効果が上がっている顧客層は？」といった自然言語での問い合わせから、瞬時に詳細な分析レポートや改善案を得る。
- 経営層・事業責任者: 「次の四半期の成長ドライバーとなる事業は何か？リスク要因は？」といった問いに対して、複数のデータソースを統合したインサイトを迅速に得る。
- データアナリスト/エンジニア: ルーチンワークから解放され、より高度な戦略的分析やデータ基盤構築に注力できる。

データ民主化、**意思決定の加速**、**生産性の向上**、**新たなビジネス価値の創出**

データ活用の「今」と「未来」	01
Google Cloud の「データ エージェント」とは？	02
「データ エージェント」がもたらすビジネス変革のシナリオ	03
データ駆動型組織への新たな一歩	04

エンド ツー エンドのデータ分析プラットフォーム

エンド ツー エンドのデータ分析プラットフォーム

THANK YOU